From: Brain, Behaviour and Evolution, D. A. Oakley & H. C. Plotkin, Eds., pp 219-228. Methuen and Company, 1979.

1979

211

9 Consciousness, freewill and personal identity[1]

R. W. Sperry

Two special properties of the brain not found in other natural systems, as far as we yet know, have always been notoriously difficult for science to deal with – even in principle. The first of these, of course, is conscious awareness, that will-o'-the wisp that science cannot find, cannot demonstrate, measure or work with and, in most cases, something just the basic nature of which we have been unable to conceive satisfactorily or even imagine. How the brain mechanisms generate subjective conscious experience continues to pose the number one problem for brain research and one of the most truly mystifying unknowns remaining in the whole of science (see also Chapters 7 and 8 of this volume).

The second brain property that science finds particularly troublesome is freewill. Science is concerned with causal relations and can hardly work out the natural laws, predictions and understanding of a system that fails to obey the principles of lawful causation. One of the earliest rules for animal behaviour stated that when rigorous conditions are established in which all sensory input can be strictly controlled, one may predict for any measured stimulus that an animal will respond 'as it damn pleases'. This was widely referred to back in the 1930s as the 'Chicago Law of Behaviour' – or, in Chicago, as the 'Harvard Law'.

It is curious and perhaps not entirely coincidental that these two brain properties that science finds so unaccountable are commonly considered by practically all of us to be the two most important and most treasured of all our brain faculties. In approaching some of the critical problems involved I turn first to related observations concerning personal identity and the unity of the self. It has been argued by Anderson (1974) that if one can have two co-conscious entities occupying the same cranium concurrently, as in commissurotomy (split-brain; see chapter 8 of this volume) subjects, and if two or more different persons can occupy the same body successively, as in multiple personality or fugue states, it follows logically that it is no longer correct to identify a 'person' or 'self' as being correlated one-to-one with a body. The concepts and definitions of 'person', 'self' and related terms need accordingly to be refined more precisely in terms of the critical brain states involved. Such definitions become important for medicolegal issues concerning, for example, comatose, anencephalic or severely deranged mental conditions, in evaluating donors for vital organ transplants, in dealing with different stages in foetal development, etc.

An interesting position in regard to the concept of 'personal identity' has been taken by Puccetti (1973, 1976) and Bogen (1969) who infer that each hemisphere must have a mind of its own – not only after brain bisection but in the normal intact state as well; a conclusion apparently accepted also by De Witt (1975) with the qualification that only the left cerebral member has self-awareness and is therefore qualified as a person. The argument goes like this: if cutting the cross-connections between the hemispheres leaves two co-conscious minds, and if surgical removal of a whole hemisphere, that is, hemispherectomy, leaves one conscious person or self, regardless of which hemisphere is removed, then there must have been two present to start with. Puccetti contends that we are, therefore, all of us, really a compound of two conscious persons that coexist in the normal brain, one based in each hemisphere, and that this goes undetected when the commissures are intact and the normally conjoined hemispheres work in perfect synchrony. A similar proposal regarding the inherent duality of mind was made back in 1844 by Wigan (see Zangwill, 1974). Again one is impressed with the need for sharpening definitions of mind, person, self and related concepts. Regardless of terminology, however, the question of whether the normal intact brain contains only one unified realm of conscious awareness or alternatively maintains

two separate conscious systems, or minds, one centred in each hemisphere, poses a rather clear dichotomy that should be subject eventually to a definite empirical answer.

My own position on this question has been a relatively conventional one. I see consciousness and the conscious self as being normally single and unified, mediated by brain processes that typically involve and span both hemispheres through the commissures. This interpretation implies, first, that the fibre systems of the brain mediate the stuff of conscious awareness as well as the switching mechanisms, synaptic interfaces or other interaction sites of the grey matter. Second, that the fibre cross-connections between the hemispheres are not different in this respect from fibre systems within each hemisphere. Third, this interpretation is based on a functionalist theory of consciousness that goes back to the early 1950s (Sperry, 1952) in which the subjective unity in conscious experience along with other subjective effects is ascribed not so much to corresponding spatio-temporal unity in neural activity or to other isomorphic or topological correspondence but rather to the operational or functional effects in brain dynamics. What counts in determining subjective meaning on these terms is the way a given brain process works in the context of cerebral organization. Subjective unity is accordingly conceived in terms of organizational and functional relations which in turn leads to the idea of a functional (thus causal) impact (see Sperry, 1976).

When I tried to put some of these threads together back in the mid-1960s, I found to my initial consternation (as well as that of immediate colleagues) that what seemed to be emerging was a conceptual formula for the way that conscious mind could move matter in the brain and exert causal influence in the direction and control of behaviour – in direct contradiction, of course, to the central founding precepts of behaviourism and of twentieth-century scientific materialism generally, and contrary to everything that we had always been taught and believed. Since the initial statement of these concepts (Sperry, 1965), however, their influence has been apparent with respect particularly to subjectivist approaches in behavioural science. As long as it remained inconceivable that phenomena of conscious experience could affect the course of brain events, those disciplines in psychology that rely on introspective reports of subjective experience, including the clinical, humanist, cognitive and related schools, continued to be put down in dominant behaviourist thinking as

something less than scientific. Once a credible conceptual model for psychophysical interaction became recognized, wherein mental phenomena as top-level controls were neither identical with nor reducible to neural events, the scientific status of consciousness and of the subjective approach underwent a change. Terms like 'mental imagery' and visual, verbal or auditory 'images' and all forms of inner thought, motivation and feeling now became more acceptable as explanatory constructs. After more than fifty years of being strictly avoided on behaviourist principles, such subjective terms have recently exploded into wide usage (Pylyshyn, 1973) in a change variously referred to as the 'cognitive' (Dember, 1974), 'humanist' or 'third' (Matson, 1973) revolution in psychology. Meantime in the mind-brain controversy, mentalists, dualists and psychophysical interactionists have suddenly begun to reappear in force, after having been essentially silent and invisible for decades.

It is not critical at this stage that the new interpretation lacks any firm proof. No proof is available, either, for the behaviourist-materialist position. Just the fact that a scientifically possible explanatory model for psychophysical interaction is conceivable has been sufficient in itself to release the long pent-up subjectivist pressures.

At the same time, more peripheral movements leaning towards the mystical and supernatural have also been bolstered secondarily in this recent mentalist upsurge, including parapsychology. Actually no direct support for these latter can be found in our present mind-brain model. If anything, the current interpretation – in which conscious experience becomes a systemic property of and functionally tied to the physical brain process (see below) – makes less likely than ever the possible occurrence of mental telepathy, psychokinesis, precognition and other so called psi phenomena. Nor can the current view be said to encourage hopes for the existence of any separate, non-physical realm of conscious mind or spirit divorced from matter. In other words the current swing away from traditional materialism does not carry us all the way back to dualistic or supernatural concepts but represents, rather, an intermediate compromise within which aspects of both classic physicalist and mentalist doctrine are fused in a new combination.

Without attempting here to review in detail these conceptual developments let me just restate briefly that, in our current interpretation of consciousness, subjective awareness is conceived to be a

functional property of neural events generated at top levels in the brain hierarchy. The emergent (functional) properties are conceived to have causal consequences in cerebral activity just as emergent properties commonly do elsewhere. The regulative control role of conscious experience is seen to be based largely in the universal power of any system as a whole over its parts. Mental phenomena built of neural events are conceived to act as dynamic entities in brain organization interacting at their own level in brain function. As high-level dynamic entities, the mental processes control their component biophysical, molecular, atomic and other subelements in the same way, for example, that the organism as a whole controls the fate of its separate organs and cells or just as the molecule as an entity carries all its component atoms, electrons and other subatomic parts through a distinctive time-space course in a chemical reaction. An expanded description of this holistic or entitative type of causal control is presented by Pols (1971).

As is the rule for part-whole relations, a mutual interaction between the neural and the mental events is indicated; the brain physiology determines the mental effects, as generally agreed, but the neurophysiology is at the same time reciprocally governed by the higher subjective properties of the enveloping mental operations, as these interact at their own level and prevail upon subsidiary events in brain dynamics. A full causal account of brain function is thus not possible in purely neurophysiological or biophysical terms that do not include these higher, yet-to-be-described mental processes with their subjective pattern properties different from the neural events *per se*.

By way of illustration, if one could render the nerve impulse and related glial activity X-ray opaque or radiant and then take fluoroscopic-like pictures of the cerebral turbulence for different kinds of conscious brain events, one might be able in time to begin to describe the critical differences that are responsible, for example, for auditory as opposed to visual or tactual sensation and later to go into further intramodal refinements describing the processing differences involved in seeing red versus green or a triangle versus a square, etc. These conscious processes, as entities, have never been described, and the objective descriptions are still far out of reach. When the objective account becomes available we will have both the objective and subjective descriptions, but the subjective effect, on these terms, should be understandable and inferable from the

objective description, because the subjective meaning depends on how the brain process, as a dynamic entity, works in the going context of brain activity (Sperry, 1952). The basic organizational features involved are assumed to be genetically determined in very large part.

The foregoing combines important features of both classic dualistic mentalism and monistic materialism. It is mentalistic in that the contents of subjective mental experience are recognized as important aspects of reality in their own right, not to be identified with the neural events as these have heretofore been conceived nor reducible to neural events. Further, the subjective mental properties and phenomena are posited to have a top-level control role as causal determinants (Sperry, 1976). On these terms mind moves matter in the brain. Not only can subjective mind no longer be ignored in science; it becomes a prime control factor in explanatory models. In former theories of consciousness at all acceptable to science, consciousness has been so defined that the causal march of brain mechanisms would proceed the same, whether it is accompanied by subjective experience or not. This is not the case in the present model.

At the same time, the current view can be called materialistic, in that the subjective phenomena as functional properties of brain activity are built of neural events and therefore always tied, as emergent properties, to the physical brain with all its anatomical and physiological constraints. The classic definitions of dualism and monism hardly apply, however, and mentalism is no longer synonymous with dualism. It has seemed preferable to describe the new position as mentalist and monist (Sperry, 1965) reserving dualism for concepts that allow for a separate existence of conscious experience apart from the functioning brain.

This mind-brain reformulation brings important logical implications also for the interpretation and outlook regarding freewill. The causal sequence of brain events leading to and determining a given voluntary act or decision no longer is conceived to be restricted to a series of neurophysicochemical activities. The emergent subjective mental properties of these physical processes, as described above, must also be taken into account and included among the controlling causal determinants. This introduces new degrees and qualities of freedom into the brain's decision-making process, lifting it above the mechanistic, physicalistic kind of determinism envisaged in classical behaviourist, stimulus-response or materialist doctrine. For

example, one's subjective desire to do this or that, along with other subjective feelings and motivations and subjective values of all kinds, plus the whole range of cognitive mental experience, may now, *per se*, influence the progression of brain events as directive causal factors. As dynamic, holistic properties, the subjective factors are not reducible to, or identifiable with, their neural constituents or as parallelistic correlates of these. In any decision to act, these conscious mental phenomena override and supersede the component physiological and biophysical events involved in the causal progression of brain activity.

A given volitional choice may depend additionally on things like the memory and the mental perspective acquired by the subject (and any consultants) over a span of many decades preceding the decision. Data from the information store of one or more libraries may be called on and funnelled into the brain code sequence that leads to the given choice. Even factors like the predicted long-term future consequences of the various alternative choices being contemplated may be included proactively in this vast vortex of cerebral factors that governs the final decision to act. Compared with the kinds of determinism that science deals with in other systems, the degree and levels of freedom in the operations of the human brain clearly set the brain and mind of man apart with the dignity of an apex post in the universe, far above all other known systems in terms of its ability to choose and to control a course of events.

Even so, one may object that this leaves our brain's decisions nevertheless all, in a sense, determined, even though at this higher, more complex mental level. We are still caught in the web of a deterministic universe and have to do what we do. Having *degrees* of freedom, in other words, does not quite make for *complete* freedom from causal control. The answer here is that complete freedom from causation would mean behaviour based purely on chance, on caprice and would result in meaningless chaos. What one wants of freewill is not to be totally freed from causation, but, rather, to have the kind of control that allows one to determine one's own actions according to one's own wishes, one's own judgement, perspective, cognitive aims, emotional desires and other mental inclinations. This, of course, is exactly what is provided in our current interpretation.

I have already stated my belief that the organizational features of the brain which give rise to conscious processes are in large part genetically determined. If this is true then consciousness must be

subject to the evolutionary process. From the standpoint of functional control, one may ask what benefits precisely are conferred by the introduction in evolution of subjective conscious effects? Thinking in regard to this question is still preliminary and speculative along lines like the following: consider the tactical difference between responding to the world directly and responding to inner conscious representations of the outside world. Wherever displacements in time or in space are advantageous, as, for example, in mental recall, in thinking and in the formation of anticipatory sets, the use of inner representations has indispensable organizational advantages (see Chapter 7 of this volume). The real world can hardly be manipulated as can inner images. Responses involving perceptual constancies in shape, size, position, etc. would seem also to be more effectively managed through the use of inner representations. Further, the employment of implicit trial-and-error responses to inner mental models and the avoidance thereby of overt response commitments, with possible errors in the real world, is a central rationale in the evolution of thinking.

The development of an inner subjective world may thus be viewed broadly as part of the evolutionary process of freeing behaviour from its initial primitive stimulus-bound condition, providing increasing degrees of freedom of choice and of originative central processing. The subjective effects have additional advantages in the driving and directing of behaviour as motivational elements and as positive and negative reinforcers. It is difficult to conceive an efficacious motivational system without subjective properties like pain, pleasure, hunger, etc. These subjective effects evolve into controlling ends in themselves in much of human behaviour.

Conscious experience may be conceived as a rather distinct entity built into brain organization and expressly designed for specific functional effects, as opposed to viewing it as a general pervasive property of complex neural integration. There is reason to believe it is present in the higher brain centres but not in the spinal cord, for example, or lower brain stem, and probably not in the cerebellum, either. The commissurotomy evidence indicates that the system for inner conscious representations in primates and cats, at least, is confined mainly to the cerebral hemispheres proper and the upper brainstem. We assume it to be rather diffusely represented within the forebrain but by no means extending throughout all neural activity at forebrain levels. On the input side of the conscious system, a

great deal of sensory processing is completed automatically and unconsciously. The integrations required for constancy effects like those for perceived position in space during head, eye and body movements, or for the union of monocular two-dimensional patterns into novel three-dimensional percepts, or the processing of elemental auditory sounds into perceived speech, etc. are extremely complex neural functions but appear to be processed without conscious mediation. Similarly, on the output side, most or all of the complicated processing required to translate conscious aims, percepts and volitional intent into appropriate motor-behaviour patterns also takes place automatically and unconsciously. The intricate arrays of requisite muscle-contraction patterns involve a complexity of neural control that goes far beyond the ability of the conscious mind to understand and direct. This is another reason to identify the conscious properties with the relatively simple holistic features rather than with the whole intricate inframechanism of brain processing.

Though representing a rather small fraction of the total brain activity in physiological terms, the conscious properties are of prime importance from the organizational standpoint. For example, the laying down, storage, cataloguing and retrieval of memories seem to proceed very largely on the basis of their holistic conscious properties rather than those of the neuronal inframechanisms. Most higher brain processing can be viewed as being designed for, and directed towards the generation, maintenance or expression of aspects of conscious awareness. Older stimulus-response and central-switchboard concepts of brain organization that arose out of spinal cord physiology and were congenial to behaviourist interpretation may be replaced by a model in which the brain is seen to be organized as a decision-making control system monitored with value priorities and in which conscious phenomena confer certain operational advantages over and above those obtainable in systems that lack consciousness.

Note

1 This chapter contains material first published in *Perspectives in Biology and Medicine* (1976) *20* (1): 9, 12–15 and *The Journal of Medicine and Philosophy* (1977) *2* (2): 115–16, 121–2, selected by the editors. We are grateful to the University of Chicago Press for permission to use extracts from the original articles.

References

Anderson, S. L. (1974) On the Problem of Personal Identity. Ph.D. dissertation, University of California, Los Angeles.

Bogen, J. E. (1969) The other side of the brain. II. An appositional mind. *Bulletin of the Los Angeles Neurological Societies 34*: 135–62.

Dember, W. N. (1974) Motivation and the cognitive revolution. *American Psychologist 29*: 161–8.

De Witt, L. (1975) Consciousness, mind and self: the implications of the split-brain studies. *British Journal for the Philosophy of Science 26*: 41–7.

Matson, F. W. (1973) Humanistic theory: the third revolution in psychology, In P. Zimbardo and C. Malach (eds) *Psychology for our Times*. Glenview, Illinois: Scott, Foresman, 19–25.

Pols, E. (1971) Power and agency. *International Philosophical Quarterly 11*: 293–313.

Puccetti, R. (1973) Brain bisection and personal identity. *British Journal for the Philosophy of Science 24*: 339–55.

Puccetti, R. (1976) The mute self: a reaction to De Witt's alternative account of the split-brain data. *British Journal for the Philosophy of Science 27*: 65–73.

Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychological Bulletin 80*: 1–24.

Sperry, R. W. (1952) Neurology and the mind-brain problem. *American Scientist 40*: 291–312.

Sperry, R. W. (1965) Mind, brain and humanist values. In J. R. Platt (ed.) *New Views on the Nature of Man*. Chicago: University of Chicago Press.

Sperry, R. W. (1976) Mental phenomena as causal determinants in brain function. In G. Globus, G. Maxwell and I. Savodnik (eds) *Consciousness and the Brain*. New York: Plenum Publishing Corp.

Zangwill, O. L. (1974) Consciousness and the cerebral hemispheres. In S. Dimond and J. Beaumont (eds) *Hemisphere Function in the Human Brain*. London: Paul Elek.